Research Article

Statistics in Medicine

Received 1 July 2012, Accepted 18 March 2013 Published online 13 May 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5821

Avoiding zero between-study variance estimates in random-effects meta-analysis

Yeojin Chung,[a*†] Sophia Rabe-Hesketh[b] and In-Hee Choi[b]

Fixed-effects meta-analysis has been criticized because the assumption of homogeneity is often unrealistic and can result in underestimation of parameter uncertainty. Random-effects meta-analysis and meta-regression are therefore typically used to accommodate explained and unexplained between-study variability. However, it is not unusual to obtain a boundary estimate of zero for the (residual) between-study standard deviation, resulting in fixed-effects estimates of the other parameters and their standard errors. To avoid such boundary estimates, we suggest using Bayes modal (BM) estimation with a gamma prior on the between-study standard deviation. When no prior information is available regarding the magnitude of the between-study standard deviation, a weakly informative default prior can be used (with shape parameter 2 and rate parameter close to 0) that produces positive estimates but does not overrule the data, leading to only a small decrease in the log likelihood from its maximum. We review the most commonly used estimation methods for meta-analysis and meta-regression including classical and Bayesian methods and apply these methods, as well as our BM estimator, to real datasets. We then perform simulations to compare BM estimation with the other methods and find that BM estimation performs well by (i) avoiding boundary estimates; (ii) having smaller root mean squared error for the between-study standard deviation; and (iii) better coverage for the overall effects than the other methods when the true model has at least a small or moderate amount of unexplained heterogeneity. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: random-effects meta-analysis; variance estimation; Bayesian posterior mode; penalized maximum likelihood

1. Introduction

Meta-analysis combines estimates of associations, or effect sizes, from several studies to estimate an overall effect size. Two common approaches are fixed-effects and random-effects meta-analysis. In fixed-effects meta-analysis, the estimates y_i of effect size from individual studies i are assumed to differ from the true overall effect size μ only because of lack of precision (sampling variability and measurement error),

$$y_i = \mu + \epsilon_i,\ i = 1, \ldots, k \tag{1}$$

where ϵ_i is the estimation error for study i with standard deviation estimated by the estimated standard error s_i of the effect size estimate y_i for study i.

The fixed-effects approach makes the strong assumption that all studies are estimating the same parameter μ. However, in medical and social research, the studies typically differ in terms of target populations (e.g., age, socioeconomic status, and severity of illness), the treatments under investigation (e.g., dosage and length of follow-up), and the outcomes being measured. Random-effects meta-analysis [1] allows for this diversity of scenarios by allowing the true effect size estimated by study i to differ from the mean effect size μ by a random amount θ_i (with mean 0 and variance τ^2):

$$y_i = \mu + \theta_i + \epsilon_i. \tag{2}$$

[a]*School of Business Administration, Kookmin University, Seoul, Korea*
[b]*Graduate School of Education, University of California, Berkeley, CA, U.S.A.*
**Correspondence to: Yeojin Chung, School of Business Administration, Kookmin University, Seoul, Korea.*
[†]*E-mail: ychung@kookmin.ac.kr*

It has been widely argued (for instance, by the National Research Council [2]) that the random-effects approach is generally more realistic and preferable to the fixed-effects approach for a number of reasons. First, the studies tend to be diverse [3, 4], whereas the fixed-effects approach assumes that they are 'functionally equivalent' [5, p.83].

Second, the variance of the random effects describes the extent of heterogeneity and is arguably equally important as the average effect size for understanding the consistency of effects across different scenarios [4, p.139]. If heterogeneity is substantial, sources of heterogeneity should ideally be investigated [6, 7]. In this case, random-effects meta-regression [8–10] can be used to explain some of the heterogeneity by using covariates while also allowing for residual heterogeneity not explained by the covariates.

Third, in the random-effects approach, μ represents the mean effect size for the population of studies with varying scenarios [5, p.83], in contrast to the fixed-effects approach where μ represents the effect size shared by a set of studies assumed to conform with one specific scenario. The former interpretation clearly has more external validity. The random-effects assumption is often believed to rely on the studies themselves being sampled from a population of studies. However, Higgins *et al.* [4] point out that it is sufficient to assume exchangeability of the study-specific effect sizes (conditional on any covariates included in the model).

Fourth, by accommodating heterogeneity due to varying scenarios, the random-effects approach allows us to make predictions about effect sizes in future scenarios. Higgins *et al.* [4] describe methods for deriving such prediction intervals and state that 'predictive distributions are potentially the most relevant and complete inferences to be drawn from a random-effects meta-analysis' (p.141). Similarly, Borenstein *et al.* [5, p.131] suggest adding a 95% prediction interval, representing the 95% range of effect sizes in future studies, to the 95% confidence interval (CI) for μ in the forest plot.

Finally, the standard error of the estimate of μ increases with the between-study variance, and the fixed-effects approach may therefore understate uncertainty regarding μ. Consequently, when the between-study variance is underestimated, coverage of CIs for the overall effect is below the nominal level [11, 12]. Draper [13, p.52] argues that allowing the study-specific effects to vary can be viewed as 'continuous model expansion' to allow for model uncertainty.

In this paper, we discuss the problem of estimating the between-study variance. Unfortunately, the variance is estimated imprecisely when the number of studies is small, and estimates of 0 can frequently occur, resulting in fixed-effects estimates of the other parameters and their standard errors. Engels *et al.* [14] reanalyzed risk differences for 125 published meta-analyses of six or more randomized clinical trials and found that boundary estimates occurred in 26% of the meta-analyses. Such boundary estimates are accompanied by non-significant tests of homogeneity, typically leading to acceptance of the null hypothesis of homogeneity. However, we argue, along with Higgins *et al.* [4, p.149], that 'such a null hypothesis is typically untenable' and agree with Borenstein *et al.* [5], Curcio and Verde, [15], Draper [13], Hardy and Thompson [16], Overton [17], Viechtbauer [18], and others that the choice of model should not be based solely on a test of homogeneity but on our understanding of whether all studies share a common effect size. Importantly, tests of homogeneity tend to have little power [16], and, as always, non-rejection of the null hypothesis cannot be seen as evidence for the null hypothesis. Curcio and Verde [15] use the profile likelihood for τ to show that a particular meta-analysis dataset is consistent with quite a large amount of heterogeneity although the maximum occurs at the boundary and argue that choosing the most-supported value of 0 is not defensible because this leads to the smallest CI for μ. Overton [17] mentions that the random effect approach may yield CIs that are too wide but argues that a conservative approach is appropriate when there is little information.

When the number of studies is small, Borenstein *et al.* [5, p.84] and Higgins *et al.* [4, p.154] suggest that a Bayesian approach with an informative prior distribution for the between-study variance may be the best option. Higgins and Whitehead [19] use an 'empirical prior' for the heterogeneity parameter that is estimated by performing a meta-analysis of historical meta-analyses. Higgins *et al.* [4, p.154] suggest that it may even be preferable to use a plausible value for the between-study variance instead of the value estimated from a small number of studies (see also Longford [20]).

In this paper, we suggest avoiding boundary estimates by specifying a Bayesian prior for the between-study variance and no priors/uniform priors for the other model parameters [21]. We estimate the

parameters by maximizing the marginal posterior of the model parameters (integrated over the random effects), which is equivalent to penalized maximum likelihood estimation. The prior is chosen to be weakly informative, keeping estimates away from the boundary and at the same time being faithful to the data.

We also discuss CIs for μ when the parameters are estimated by maximum likelihood. The standard method for estimating CIs in this case is based on the *expected* information matrix and ignores uncertainty regarding the between-study variance τ^2. We suggest using the *observed* information matrix, which can be viewed as constructing a CI by using a quadratic approximation to the profile log likelihood of μ and hence takes uncertainty regarding τ^2 into account.

The outline of the paper is as follows. In Section 2, we illustrate the problem of boundary estimates by using two real datasets. Sections 3 and 4 review frequentist and Bayesian estimation methods for random-effects meta-analysis. Section 5 introduces our Bayes modal (BM) approach for meta-analysis without covariates, and Section 6 briefly discusses the different estimators for meta-regression with covariates. In Section 7, the different estimators are applied to the examples introduced in Section 2, and in Section 8, they are compared using simulations. We conclude with a brief discussion in Section 9.

2. Examples

2.1. Exercise for depression data

We demonstrate the boundary estimation problem by using data on 10 randomized clinical trials of exercise as an intervention in the management of depression [22, 23]. The estimated effect sizes y_i are standardized mean differences in depression scores between the treatment and control groups, and these have similar estimated standard errors between 0.24 and 0.57. We estimate a meta-regression model with two covariates: 'abstract' x_{1i}, a dummy variable for publication being a conference abstract (versus journal article or other) and 'duration' x_{2i}, the (mean-centered) length of follow-up in the number of weeks. The model can be written as

$$y_i = \mu + \beta_1 x_{1i} + \beta_2 x_{2i} + \theta_i + \epsilon_i,\ i = 1, \ldots, 10$$

where $\theta_i \sim N(0, \tau^2)$ and $\epsilon_i \sim N\left(0, s_i^2\right)$. The standard random-effects meta-analysis model corresponds to this model with $\beta_1 = \beta_2 = 0$.

The model with and without covariates was estimated by maximum likelihood (ML) and restricted (or residualized) maximum likelihood (REML [24]) by using the `metaan` [25] and `metareg` [26] commands in Stata. In the meta-analysis model without covariates, the ML estimate of μ is -1.056 with standard error 0.234 (Table I). The REML estimate is also -1.056 with a slightly larger standard error of 0.248. The between-study standard deviation τ is estimated as 0.621 and 0.671 by using ML and REML,

Table I. Parameter estimates for the meta-analysis and the meta-regression models applied to the exercise for depression dataset.

	Coefficient estimates				
	$\hat{\mu}$ (se($\hat{\mu}$))	$\hat{\beta}_1$ (se($\hat{\beta}_1$))	$\hat{\beta}_2$ (se($\hat{\beta}_2$))	$\hat{\tau}$	Log likelihood
Meta-analysis model					
DL	-1.056 (0.239)			0.640	-11.326
UMM	-1.056 (0.251)			0.683	-11.365
ML	-1.056 (0.234)			0.621	-11.321
REML	-1.056 (0.248)			0.671	-11.350
BM	-1.056 (0.253)			0.689	-11.372
Meta-regression model					
ML, REML	-0.866 (0.137)	-1.244 (0.341)	0.121 (0.053)	0 (0.162*)	-2.243
BM	-0.852 (0.153)	-1.255 (0.371)	0.122 (0.060)	0.168 (0.124*)	-2.766

*Standard error (se) of $\hat{\tau}$ based on the observed information.

DL, DerSimonian and Laird; UMM, unweighted method of moment; ML, maximum likelihood; REML, restricted maximum likelihood; BM, Bayes modal.

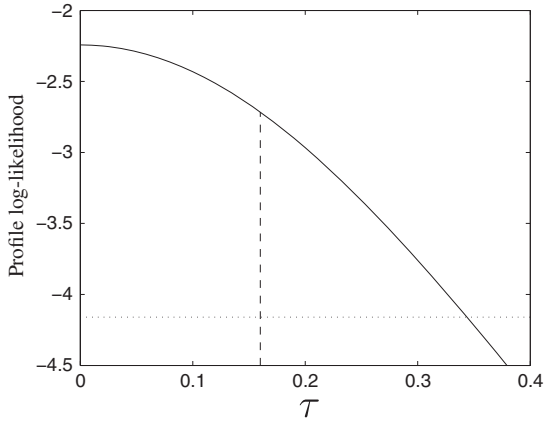

Figure 1. Profile log-likelihood function of τ for the meta-regression model applied to the exercise for depression dataset.

respectively, suggesting that there is important variability between studies. However, when we include the covariates, the residual between-study standard deviation τ is estimated as 0. This point estimate implies that the study-specific effects are perfectly predicted by the two covariates, which is unrealistic. The estimates of the regression coefficients using ML or REML are the same as for fixed-effects meta-regression, namely $\hat{\beta}_1$ and $\hat{\beta}_2$, which are -1.244 and 0.121, respectively, with estimated standard errors of 0.341 and 0.053. Both coefficients are statistically significant at the 5% level. However, note that the estimated standard errors are based on $\hat{\tau} = 0$ and are smaller than for any $\hat{\tau} > 0$, resulting in the most anti-conservative inferences.

Figure 1 shows the profile log likelihood for τ for the meta-regression model (maximized with respect to μ, β_1, and β_2). Clearly, the maximum of the profile log likelihood is attained at $\tau = 0$ but for slightly larger values of τ, e.g., $\tau = 0.162$ (=standard error of ML estimate $\hat{\tau}$, marked with the vertical dashed line), the profile log likelihood does not decrease substantially. Comparing the change in profile log likelihood with half of the 0.95 quantile of a chi-square distribution with one degree of freedom ($\chi^2_{0.95}(1)/2 = 1.921$), we can infer that values for τ up to about 0.35 are reasonably supported by the data (See the horizontal dotted line.). When τ is set to 0.35, the estimated standard error for $\hat{\beta}_1$ increases from 0.341 to 0.454, and the estimated standard error for $\hat{\beta}_2$ increases from 0.053 to 0.075. The coefficient of duration is no longer statistically significant at the 5% level. The width of the estimated prediction interval for the effect size of a new study [5, p.129] also increases with $\hat{\tau}$ and is hence understated if τ is underestimated.

2.2. *Antiplatelet therapy for prevention of stroke*

As another example, we consider data from a collaborative meta-analysis of randomized trials of antiplatelet therapy for the prevention of death, myocardial infarction, and stroke in high-risk patients [27]. Specifically, we consider nonfatal strokes in the six trials that compared dipyridamole plus aspirin with aspirin alone for patients with previous stroke, transient ischaemic attack, or myocardial infarction. Analyzing the log odds ratios, we found that the ML and REML estimates of the between-study variance are both 0. One large study with more than twice as many patients as the next largest study and with a high incidence of nonfatal stroke has a particularly small standard error and contributes 66% to the ML estimate of the mean log odds ratio $\widehat{\mu}$ in the meta-analysis model. The corresponding estimated odds ratio is 0.728, and its estimated 95% CI is (0.593, 0.896).

However, as in the previous example, the profile log likelihood for τ is relatively flat, and values of τ as large as 0.415 are supported (the profile log likelihood decreases by $\chi^2_{0.95}(1)/2$). When $\hat{\tau} = 0.415$, the contribution of the largest study to the estimated log odds ratio is only 33%, and the estimated odds ratio is 0.791 (with 95% CI from 0.484 to 1.293). The point estimate of μ is sensitive to changes in $\hat{\tau}$ because the studies differ considerably in their sizes and their relative weighting is greatly affected by the value of $\hat{\tau}$. The width of the estimated CI also increases with $\hat{\tau}$.

3. Classical estimation methods for random-effects meta-analysis

3.1. Estimation of τ^2

One of the most widely used methods for estimating τ^2 is the method of moment estimator by DerSimonian and Laird (DL [1]), which equates $\sum s_i^{-2}(y_i - \hat{\mu})^2$ with its expectation where $\hat{\mu}$ is defined as

$$\hat{\mu} = \frac{\sum_i y_i/s_i^2}{\sum_i 1/s_i^2}. \tag{3}$$

The resulting estimator is given by

$$\hat{\tau}^2_{\mathrm{DL}} = \frac{\sum_i s_i^{-2}(y_i - \hat{\mu})^2 - (n-1)}{\sum_i s_i^{-2} - \sum_i s_i^{-4}/\sum_i s_i^{-2}}. \tag{4}$$

For this method, distributional assumptions for θ_i and ϵ_i are not required. This estimator is also called the weighted method of moment estimator because it equates the weighted sum of squares with its expectation.

The unweighted method of moment (UMM) estimator equates the unweighted sum $\sum_i (y_i - \bar{y})^2$ with its expected value and solves for τ^2, resulting in

$$\hat{\tau}^2_{\mathrm{UMM}} = \frac{1}{k-1}\sum_{i=1}^{k}(y_i - \bar{y})^2 - \frac{1}{k}\sum_i s_i^2.$$

This estimator is also known as Hedges estimator [28] and coincides with the minimum norm quadratic unbiased estimator (MINQUE [29, 30]).

Under the assumptions that θ_i and ϵ_i are normally distributed, we can estimate τ^2 by ML. The log-likelihood function is given by

$$l(\mu, \tau^2) = -\frac{1}{2}\sum_{i=1}^{k}\left[\log\left[2\pi\left(s_i^2 + \tau^2\right)\right] + \frac{(y_i - \mu)^2}{s_i^2 + \tau^2}\right] \tag{5}$$

and the ML estimator is

$$\widehat{\tau}^2_{\mathrm{ML}} = \left(\frac{\sum_i \widehat{w}_i^2\left[(y_i - \widehat{\mu}_{\mathrm{ML}})^2 - s_i^2\right]}{\sum_i \widehat{w}_i^2}\right)^+,$$

where $(\cdot)^+ = \max(\cdot, 0)$ and

$$\widehat{\mu}_{\mathrm{ML}} = \frac{\sum_i \widehat{w}_i y_i}{\sum_i \widehat{w}_i}, \quad \widehat{w}_i = \left(s_i^2 + \widehat{\tau}^2_{\mathrm{ML}}\right)^{-1}. \tag{6}$$

Because these equations do not have a closed form solution, $\hat{\tau}^2_{\mathrm{ML}}$ and $\hat{\mu}_{\mathrm{ML}}$ have to be found iteratively.

It is well known that the ML estimator for variance components is downward biased in finite samples. Therefore, REML, which takes the degrees of freedom into account, is often used, particularly for small sample sizes. When $s_i = s$ for all i, unlike the ML estimator, the REML estimator of τ^2 is unbiased if it is allowed to be negative. For the random-effects meta-analysis model, the residualized log-likelihood function becomes

$$l_{\mathrm{R}}(\mu, \tau^2) = l(\mu, \tau^2) - \frac{1}{2}\log\sum_i\left(\tau^2 + s_i^2\right)^{-1}. \tag{7}$$

The REML estimator of τ^2 is given by

$$\widehat{\tau}^2_{\mathrm{REML}} = \left(\frac{\sum_i \widehat{w}_i^2\left[(y_i - \widehat{\mu}_{\mathrm{REML}})^2 - s_i^2\right]}{\sum_i \widehat{w}_i^2} + \frac{1}{\sum_i \widehat{w}_i}\right)^+$$

where $\hat{\mu}_{\mathrm{REML}}$ is the same as (6) except that $\hat{\tau}^2_{\mathrm{ML}}$ is replaced with $\hat{\tau}^2_{\mathrm{REML}}$. Morris [31] suggests an approximate REML estimator in the empirical Bayes framework, which requires an iterative method.

3.2. Estimation of μ

With known variance τ^2, the generalized least squares estimator of μ for model (2) is

$$\hat{\mu} = \frac{\sum_i y_i \left(\tau^2 + s_i^2\right)^{-1}}{\sum_i \left(\tau^2 + s_i^2\right)^{-1}}. \tag{8}$$

This estimator is unbiased, and its variance,

$$\text{var}(\hat{\mu}) = \frac{1}{\sum_i \left(\tau^2 + s_i^2\right)^{-1}}, \tag{9}$$

attains the minimum among all linear unbiased estimators by the Gauss–Markov theorem.

The fixed-effects model in (1) assumes that $\tau^2 = 0$ and the estimator of μ reduces to the weighted average in (3). In contrast, if we assume that the variance of the random effect is infinite, then $\hat{\mu}$ becomes the simple average $\bar{y} = \sum_i y_i / k$. In practice, an estimate, such as $\hat{\tau}^2_{\text{DL}}$ or $\hat{\tau}^2_{\text{UMM}}$, is substituted for τ^2 in (8), or, for ML and REML, μ is estimated iteratively with τ^2.

Asymptotically, the ML estimator $\hat{\mu}_{\text{ML}}$ follows a normal distribution with mean μ and variance given in (9). In addition, using any of the consistent estimators for τ^2 discussed in the previous section results in estimators for μ that are asymptotically equivalent to $\hat{\mu}_{\text{ML}}$ and therefore asymptotically efficient.

3.3. Methods of constructing confidence interval of μ

Asymptotically, the ML estimator $(\hat{\mu}_{\text{ML}}, \hat{\tau}^2_{\text{ML}})$ follows a normal distribution with mean (μ, τ^2) and covariance matrix given by the inverse of the information matrix. The information matrix $\mathcal{I}$ is the expectation of the negative Hessian matrix H of the log-likelihood function, given by

$$\mathcal{I} = E[-H] = \begin{bmatrix} \sum_i \left(s_i^2 + \tau^2\right)^{-1} & 0 \\ 0 & \frac{1}{2}\sum_i \left(s_i^2 + \tau^2\right)^{-2} \end{bmatrix}.$$

Because the information matrix is diagonal, its inverse is the matrix with reciprocals of each diagonal element. On the basis of the asymptotic normal distribution of $\hat{\mu}_{\text{ML}}$ and $\hat{\mu}$ with the other estimators of τ^2 (which are asymptotically equivalent to $\hat{\mu}_{\text{ML}}$), we can construct a $(1-\alpha)100\%$ Wald-type CI with lower and upper bounds given by

$$\hat{\mu} \pm z_{1-\alpha/2} \left(\sum_i \left(s_i^2 + \tau^2\right)^{-1} \right)^{-\frac{1}{2}}, \tag{10}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution. Because the true τ^2 is unknown, we replace it with an estimate. Most standard statistical software (e.g., the `metafor` package in R [32] and `metaan` in Stata) estimates CIs for μ in this way.

Because the information matrix is diagonal, we can see that the variance of $\hat{\mu}$ does not take into account the uncertainty regarding τ^2. Hardy and Thompson [33] therefore suggest constructing CIs by using the profile log likelihood for μ (after profiling out τ^2), which can be written as

$$l_{\text{p}}(\mu) = -\frac{1}{2}\sum_i \left[\log\left[2\pi\left(s_i^2 + \tau^2(\mu)\right)\right] + \frac{(y_i - \mu)^2}{s_i^2 + \tau^2(\mu)} \right],$$

where $\tau^2(\mu)$ maximizes the log likelihood in (5) for given μ. The lower and upper bounds of the CI for μ are found by the roots of

$$l_{\text{p}}(\mu) = l_{\max} - 0.5\chi^2_{1-\alpha}(1), \tag{11}$$

where $\chi^2_{1-\alpha}(1)$ is $(1-\alpha)$th quantile of the χ^2 distribution with one degree of freedom and $l_{\max}$ is the maximum of the likelihood function. Because the profile log-likelihood function is not always symmetric in μ, the profile likelihood CI is not necessarily centered at $\hat{\mu}$ unlike the Wald-type CI in (10).

An alternative to the profile likelihood approach that also takes into account parameter uncertainty for τ^2 is to construct a Wald-type interval based in the *observed* information [34, 35], defined as minus the Hessian,

$$I(\mu,\tau^2) = -H = \begin{bmatrix} \sum_i \frac{1}{s_i^2+\tau^2} & \sum_i \frac{y_i-\mu}{(s_i^2+\tau^2)^2} \\ \sum_i \frac{y_i-\mu}{(s_i^2+\tau^2)^2} & \frac{1}{2}\sum_i \frac{2(y_i-\mu)^2}{(s_i^2+\tau^2)^3} - \frac{1}{(s_i^2+\tau^2)^2} \end{bmatrix},$$

evaluated at the estimates $(\hat{\mu},\hat{\tau}^2)$. This approach is equivalent to approximating the profile log-likelihood function by a quadratic function at the mode and then finding the roots of (11) for this quadratic approximation [36, p.267].

Different from $\mathcal{I}$, $I(\hat{\mu},\hat{\tau}^2)$ is not always diagonal, and, generally, $(I^{-1})_{1,1} \neq (I_{1,1})^{-1}$. The variance of $\hat{\mu}$, on the basis of the observed information, is

$$(I^{-1})_{1,1} = \left[\left(\sum_i \frac{1}{s_i^2+\hat{\tau}^2}\right) - \left(\sum_i \frac{y_i-\hat{\mu}}{\left(s_i^2+\hat{\tau}^2\right)^2}\right)^2 \Big/ \left(\sum_i \frac{(y_i-\hat{\mu})^2}{\left(s_i^2+\hat{\tau}^2\right)^3} - \frac{1}{2}\sum_i \frac{1}{\left(s_i^2+\hat{\tau}^2\right)^2}\right)\right]^{-1}. \tag{12}$$

Comparing the (1,1)-element of $\mathcal{I}^{-1}$ with (12), we find that the variance estimate based on the observed information is always greater than or equal to the variance estimate based on the expected information. Equality holds when $\hat{\mu} = \left(\sum_i y_i/\left(s_i^2+\hat{\tau}^2\right)^2\right) / \left(\sum_i \left(s_i^2+\hat{\tau}^2\right)^{-2}\right)$, which includes the special case $s_i = s$ for all i.

4. Bayesian estimation method using Markov chain Monte Carlo

Because of advances in computational methods, fully Bayesian estimation via Markov chain Monte Carlo (MCMC) has become popular. The Bayesian random-effects meta-analysis model can be written as

$$y_i|\theta_i \sim N\left(\theta_i, s_i^2\right),\ \theta_i|\mu,\tau \sim N(\mu,\tau^2),\ \mu \sim p(\mu),\ \tau \sim p(\tau).$$

A diffuse normal prior distribution is often used for μ, for example, $N(0, 10^6)$. Unlike the prior for μ, the choice of prior distribution for τ can have a large impact on statistical inference especially when the number of studies is small [37]. Therefore, a range of different seemingly vague priors has been used for the between-study variance.

An inverse-gamma prior on τ^2 with small values of the hyperparameters is a popular choice [38–41] that is mathematically convenient because of its conjugacy. As an alternative, a uniform prior on τ, for example with range $(0, 100)$, is also often used [4, 42].

Lambert *et al.* [37] compare 13 vague priors for τ (or τ^2), focusing on their performance when the number of studies is small. Their extensive simulation comparison of point estimates (posterior median of τ and μ) and coverage probabilities shows that the inverse-gamma$(10^{-3}, 10^{-3})$ on τ^2 and the uniform $(0, 100)$ on τ behave better than the other priors in terms of bias, coverage, and convergence problem.

5. Bayes modal estimator with a weakly informative prior

In this paper, we suggest specifying a prior $p(\tau)$ for τ and maximizing the resulting posterior distribution to avoid boundary estimates. Such posterior modal estimation has been used to avoid boundary estimates in log-linear models [43] and latent class analysis [44, 45] and to obtain more stable estimates of item parameters in item response theory [46–48].

We suggest specifying a gamma prior on the between-group standard deviation (see Chung *et al.* [21] for a similar approach in linear random-intercept models for clustered data). With the gamma(α,λ) prior on τ, $p(\tau) = \frac{1}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\lambda\tau}$, and implicitly assuming an improper uniform prior for μ, we can write the log-posterior of (μ,τ) as

$$p(\mu,\tau|\boldsymbol{y}) = l(\mu,\tau^2) + (\alpha-1)\log\tau - \lambda\tau + c_1, \tag{13}$$

where $l(\mu,\tau^2)$ is the log likelihood in (5). The parameter estimate for (μ,τ) that maximizes $p(\mu,\tau|\boldsymbol{y})$ can be found numerically. Unlike posterior mean (or median) estimation, the posterior mode can be

found by numerical optimization as for ML estimation and does not require simulation. Because we consider a prior for τ only, the BM estimator for μ will have the same form as the ML estimator in (6) but with different weights because of the different estimates of τ. Because the gamma density with $\alpha > 1$ is 0 at the origin, the posterior density for τ cannot have a mode at the boundary even if the likelihood function is maximized at the boundary. In addition, with $\alpha = 2$, the gamma density increases linearly around 0, which keeps us from assigning excessive penalty near the boundary.

The effect of the gamma penalty can be examined using a quadratic approximation of the profile log-likelihood function as a function of τ (not τ^2) after profiling out μ. The profile log-likelihood function of τ is

$$l_{\mathrm{p}}(\tau) = -\frac{1}{2}\sum_i \left[\log\left[2\pi\left(s_i^2+\tau^2\right)\right] + \frac{(y_i-\mu(\tau))^2}{s_i^2+\tau^2},\right]$$

where $\mu(\tau) = \sum_i y_i\left(\tau^2+s_i^2\right)^{-1} / \sum_i \left(\tau^2+s_i^2\right)^{-1}$. The first derivative of $l_{\mathrm{p}}(\tau)$ is

$$\begin{aligned}\frac{\partial l_{\mathrm{p}}(\tau)}{\partial \tau} &= (2\tau)\cdot\frac{\partial l_{\mathrm{p}}(\tau)}{\partial(\tau^2)}\\ &= (2\tau)\cdot\sum_i -\frac{1}{2}\left[\frac{1+2(y_i-\mu(\tau))\frac{\partial\mu}{\partial(\tau^2)}}{s_i^2+\tau^2} - \frac{(y_i-\mu(\tau))^2}{\left(s_i^2+\tau^2\right)^2}\right].\end{aligned} \tag{14}$$

The summation in (14) represents $\partial l_{\mathrm{p}}/\partial(\tau^2)$, which is possibly negative when the maximum of l_{p} (as a function of τ^2 with restriction $\tau^2 \geqslant 0$) is attained at $\tau^2 = \tau = 0$. Even in this case, $\partial l_{\mathrm{p}}/\partial\tau$ is 0 at the boundary $\tau = 0$. Therefore, the leading term of the Taylor expansion of $l_{\mathrm{p}}(\tau)$ at the ML estimate of τ is quadratic.

Consequently, we consider the quadratic approximation of the profile log likelihood around the ML estimator, $\hat{\tau}_{\mathrm{ML}}$, as follows:

$$l_{\mathrm{p}}(\tau) \approx -\frac{(\tau-\hat{\tau}_{\mathrm{ML}})^2}{2\cdot \mathrm{se}(\hat{\tau}_{\mathrm{ML}})^2} + c_2.$$

With the gamma(α,λ) prior on τ, the profile log-posterior density (where μ is profiled out) can be approximated by

$$\log p_{\mathrm{p}}(\tau|\boldsymbol{y}) \approx -\frac{(\tau-\hat{\tau}_{\mathrm{ML}})^2}{2\cdot \mathrm{se}(\hat{\tau}_{\mathrm{ML}})^2} + (\alpha-1)\log\tau - \lambda\tau + c_3. \tag{15}$$

For a fixed value of α, the mode of the right hand side of (15) increases as λ decreases, and so the largest posterior mode is obtained when $\lambda \to 0$. When $\hat{\tau}_{\mathrm{ML}} = 0$, the largest BM estimator is $\hat{\tau}_{\mathrm{BM}} = \mathrm{se}(\hat{\tau}_{\mathrm{ML}})\sqrt{\alpha-1}$ with $\lambda \to 0$ [21]. Therefore, if we use gamma$(2,\lambda)$ with $\lambda \to 0$, the BM estimator is only about one standard error away from the ML estimator when the ML estimator is on the boundary. This also implies that the log likelihood does not decrease considerably from the maximum when we use the BM estimator.

Therefore, as a default choice of (α,λ), we select $\alpha = 2$ and $\lambda \approx 0$, but if one has stronger prior information about τ, $\alpha > 2$ can be used. Although this choice assigns more penalty on the boundary, this is acceptable if it is consistent with the prior information. By observing that the gamma(α,λ) prior has its mode at $(\alpha-1)/\lambda$, we can then set λ to $(\alpha-1)/\tau^*$ where τ^* is a plausible value of τ on the basis of prior knowledge.

In terms of suggesting a default prior, our approach can be regarded as an objective Bayesian method [49]. At the same time, from the frequentist point of view, BM estimation can be viewed as penalized ML estimation with an additive penalty term $(\alpha-1)\log\tau+\lambda\tau$. From this view, our method can be compared with REML, which maximizes l_{R} in (7). Comparing the REML penalty term $-0.5\log\sum_i\left(\tau^2+s_i^2\right)^{-1}$ in (7) with the log-gamma$(2,\lambda)$ with $\lambda \approx 0$, we can see that the terms differ by a constant if s_i^2 is negligible compared with τ^2.

Figure 2 compares the REML penalty term for three different sample sizes ($k = 5, 10, 30$) with the log-gamma penalty term. The within-study error variance s_i^2 is set to 0.1 for all i, and the intra-class correlation $\rho = \tau^2/(\tau^2+s^2)$ is shown. As τ increases, all the penalty functions become parallel. We can

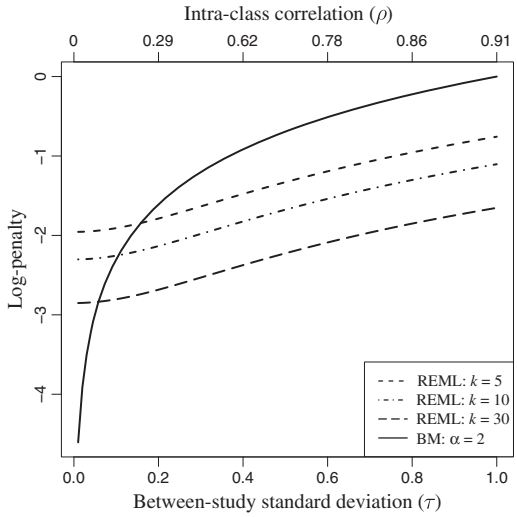

Figure 2. Log-penalty (log-prior) functions for restricted maximum likelihood (REML) and Bayes modal (BM) using gamma$(2, 0)$ with $s_i = 0.1$ for all i. Intra-class correlation $\rho = \tau^2/(\tau^2 + s^2)$ is reported on the second x-axis.

infer that the REML and the BM estimators are close when τ is large compared with s_i^2, with an intra-class correlation of about 0.6 or more. For smaller values of τ, the gamma prior assigns more penalty than REML and does not permit boundary estimates of τ.

6. Meta-regression model

When study-level covariates are available, as in the exercise for depression data, a meta-regression model may be appropriate. As an extension of the random-effects meta-analysis model in (2), the meta-regression model with m covariates can be written as

$$y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \theta_i + \epsilon_i, \quad i = 1, \ldots, k, \tag{16}$$

where $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{im})^T$ is a vector of covariates for study i, $\boldsymbol{\beta} \in \mathcal{R}^{m+1}$ is a vector of fixed coefficients, $\theta_i \sim N(0, \tau^2)$, and $\epsilon_i \sim N\left(0, s_i^2\right)$.

The log-likelihood function for model (16) is

$$l(\boldsymbol{\beta}, \tau^2) = \frac{1}{2} \sum_i \left[\log\left(\tau^2 + s_i^2\right) + \frac{(y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i)^2}{\tau^2 + s_i^2} \right]. \tag{17}$$

As in the meta-analysis model with no covariates, the ML estimator does not have a closed form but must be found iteratively. On the basis of the information matrix, the asymptotic variance of the ML estimator for $\boldsymbol{\beta}$ is

$$\text{var}(\hat{\boldsymbol{\beta}}_{\text{ML}}) = \left(\sum_i \frac{1}{\tau^2 + s_i^2} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1}$$

and this variance is estimated by replacing τ^2 with its estimate. Therefore, when τ^2 is underestimated, the uncertainty of the fixed coefficients will be underestimated, leading to inflated probabilities of type-I errors.

For BM estimation with a gamma prior on τ, the log-posterior function has the same form as in (13) but with the log likelihood replaced by (17). For comparison, the restricted log-likelihood function for the meta-regression model is given by

$$l_{\text{R}}(\boldsymbol{\beta}, \tau^2) = l(\boldsymbol{\beta}, \tau^2) - \frac{1}{2} \log \left| \sum_i \frac{1}{\tau^2 + s_i^2} \boldsymbol{x}_i \boldsymbol{x}_i^T \right|. \tag{18}$$

When $s_i^2 = s^2$ for all i and x_{ij} is 1 only for $i = j$ and 0 otherwise, the REML-penalty term (the second term in the right-hand side of (18)) reduces to

$$\frac{m+1}{2}\log(\tau^2 + s^2) + \mathrm{c},$$

where c is a constant. Compared with the log-gamma penalty, $(\alpha - 1)\log\tau + \lambda\tau$, we observe that the choices of $\alpha = m + 2$ and $\lambda \approx 0$ make the gamma penalty similar to the REML penalty when s^2 is small. The gamma prior tends to assign more penalty around $\tau = 0$ than the REML penalty, as shown in Figure 2 for the meta-analysis model. As s^2 decreases, the REML penalty approaches the gamma penalty up to an additive constant.

The meta-regression model is equivalent to the Fay–Herriot model (1979) for small area estimations, and for the latter, adjustment for density maximization [51, 52, 53] has been proposed to obtain a strictly positive group-level variance estimate. Morris and Tang use a penalty term $\pi(\tau^2) = (\tau^2)^{c-1}$, which is equivalent to our BM estimator with a gamma(α, λ) prior on τ, with $\alpha = 2c + 1$ and $\lambda \to 0$. Therefore, our BM estimator also shares the properties of adjustment for density maximization, such as predictions of θ_i being minimax for mean squared-error loss when the within-group variances are equal and c is equal to one or some smaller values [53].

7. Examples revisited

For the data described in Section 2, estimates of μ, β_1, β_2, and τ by using BM estimation with a gamma$(2, 10^{-4})$ prior on τ were obtained using the Stata program `gllamm` [54, 55]. For the depression data, BM estimates are given in Table I together with DL, UMM, ML, and REML estimates. DL, ML, and REML estimates were obtained using `metaan` and `metareg`, and we wrote our own program to obtain UMM estimates. For the meta-analysis model, all the methods give the same estimate for μ, but for τ, the BM estimate is the largest, and the ML estimate is the smallest. The log likelihoods are all very similar.

For the meta-regression model applied to the depression data, the BM estimate of τ is 0.168, whereas the ML and REML estimates are both 0. As expected, the BM estimate of τ is about one standard error (0.162) away from the ML estimate of 0, and the log likelihood decreases only 0.523 from the maximum value of -2.243, which shows that the BM estimate $\hat{\tau}_{\mathrm{BM}} = 0.168$ is reasonably supported by the data. As expected, the estimated standard errors of $\hat{\mu}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are larger for BM than for ML and REML.

For the antiplatelet therapy data, the BM estimate of τ is 0.188, which is a bit more than one standard error (0.141) away from the ML estimate of 0. The associated decrease in the log likelihood is only 0.631. Whereas the estimated fixed coefficients for the depression data (which are more balanced) are close for ML, REML, and BM methods, the estimated odds ratio for antiplatelet therapy data differs substantially between ML and REML (0.728), and BM (0.766). This shows that, if the sizes of the studies (or the standard errors of effect sizes) differ greatly, BM can have large influence on the point estimate and standard error estimate for μ because the relative weighting of the studies is affected by $\hat{\tau}$. The estimated 95% CI for the odds ratio by BM is (0.556,1.057), which is wider than the CI by ML (0.593,0.896).

We now compare and visualize the different types of CIs for μ discussed in Section 3.3 by using the depression data. Consider the likelihood surface for the meta-analysis model without covariates shown as a contour plot in Figure 3. The 'X' marks the ML estimates $(\hat{\mu}_{\mathrm{ML}}, \hat{\tau}_{\mathrm{ML}}) = (-1.056, 0.689)$. The dashed curve shows the values of τ that maximize the log likelihood for given values of μ. The log likelihood along this path, as a function of μ, is the profile log-likelihood function $l_{\mathrm{p}}(\mu)$. The dash-dotted line indicates where the conditional log likelihood given the ML estimate of τ is located. The contours show decreases from the maximum (-11.32) in multiples of $\chi^2_{0.95}(1)/2 = 1.92$.

Figure 4 illustrates four different methods for constructing CIs for μ. The top panel shows the conditional log likelihood given $\tau = \hat{\tau}_{\mathrm{ML}}$, which is a two-dimensional view of Figure 3 along the dash-dotted line. Note that, for any given τ, $\hat{\mu}_{\mathrm{ML}}$ is a linear combination of the y_is, which are assumed to follow normal distributions, and so $\hat{\mu}_{\mathrm{ML}}$ is also exactly normally distributed. Therefore, the conditional log-likelihood function of μ given τ is always quadratic in μ, as observed in (5) and in the top panel of Figure 4. The vertical solid line is at $\hat{\mu}_{\mathrm{ML}} = -1.056$, and the two dashed lines indicate the lower (-1.515) and upper (-0.597) bounds of the 95% Wald-type CI of μ on the basis of the expected information in (10) at $(\hat{\mu}_{\mathrm{ML}}, \hat{\tau}_{\mathrm{ML}})$. The horizontal solid line is 1.92 $(= \chi^2_{0.95}(1)/2)$ lower than the maximum so that it crosses at the bounds of the CI.

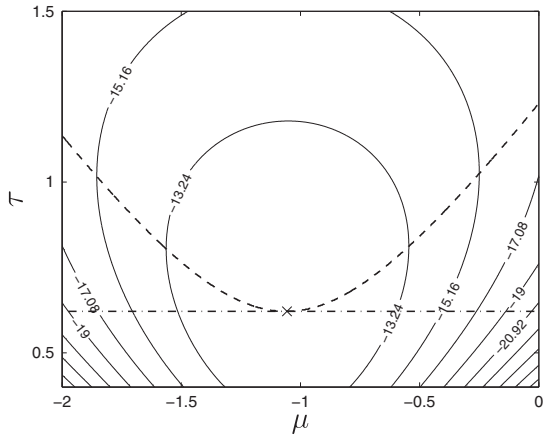

Figure 3. Contour plot of the log-likelihood function for the meta-analysis without covariates applied to the exercise for depression dataset.

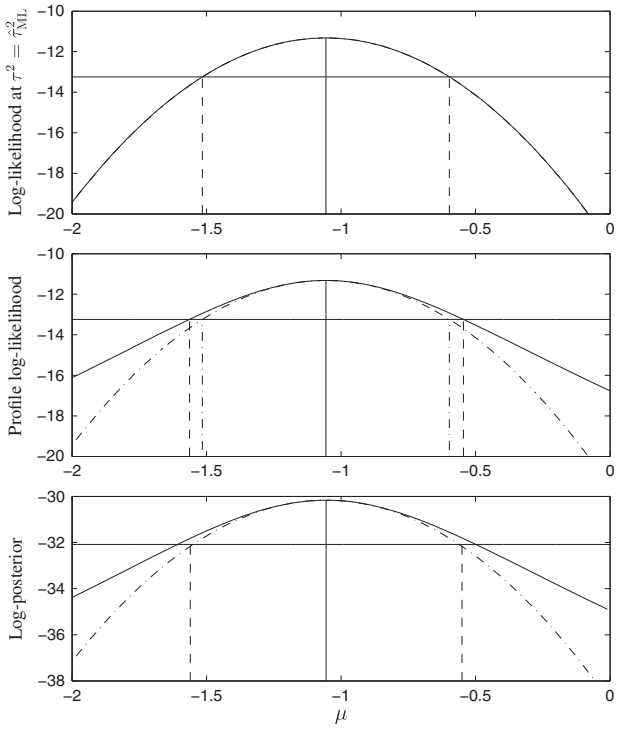

Figure 4. Comparison of the Wald-type confidence interval (CI) based on the expected information with maximum likelihood (ML) estimates (top), the profile likelihood CI and Wald-type CI based on the observed information with ML estimates (middle), and the Wald-type CI based on Bayes modal estimation (bottom) for the meta-analysis applied to the exercise for depression dataset.

The plot in the middle panel of Figure 4 shows the profile log-likelihood $l_{\mathrm{p}}(\mu)$. Again, the solid horizontal line is 1.92 lower than the maximum of the profile log-likelihood function. Therefore, the limits of the 95% profile likelihood CI $(-1.563, -0.545)$ are where the profile log-likelihood curve and the horizontal line cross. Because the profile log likelihood is not always symmetric, the CI might not be centered at $\hat{\mu}_{\mathrm{ML}}$. The dash-dotted curve is a quadratic approximation of the profile-likelihood (PL) function at the mode. The curvature at the mode can be calculated from the observed information at $(\hat{\mu}_{\mathrm{ML}}, \hat{\tau}_{\mathrm{ML}})$. The limits of the 95% Wald-type CI $(-1.515, -0.597)$ based on the observed information are shown as vertical dash-dotted lines, which coincide with the points where the quadratic curve and the horizontal line cross each other. In this dataset, the Wald-type CIs based on the expected and observed

information are almost the same, but generally, the latter will be wider because it takes into account the uncertainty of $\hat{\tau}$.

The bottom panel shows the log-posterior at $\tau = \hat{\tau}_{\mathrm{BM}}$ with a gamma$(2, 10^{-4})$ prior on τ. The limits of the 95% Wald-type CI $(-1.551, -0.561)$ based on the 'observed information' (the Hessian of the log-posterior function) are shown as dashed vertical lines. This CI is close to the profile likelihood CI but is the widest among all the intervals considered here.

8. Simulation study

8.1. Simulation design

To compare the BM estimators with the alternatives discussed in Section 3, we conduct simulations for the meta-analysis model with no covariates and the meta-regression model with two covariates. The data for the meta-analysis model with no covariates are simulated for different numbers of studies ($k = 5$, 10, 30) and for different values of the between-study variance ($\tau^2 = 0$, 0.01, 0.05, 0.1, 0.2). Although the BM estimator is based on the assumption that $\tau^2 > 0$, we include $\tau^2 = 0$ here as a worst-case scenario for the BM estimator. For each combination of k and τ^2, we generate 1000 datasets with true overall effect $\mu = 0.5$ and heterogeneous within-study variances s_i^2. We follow Brockwell and Gordon [11] by drawing within-study variances from a $0.25\chi^2(1)$ distribution truncated below at 0.009 and above at 0.6. The mean and standard deviation of this distribution of s_i^2 are 0.20 and 0.21, and the intra-class correlations computed by plugging in the mean of s_i^2 are $\rho = 0$, 0.05, 0.20, 0.33, and 0.5, respectively, for $\tau^2 = 0$, 0.01, 0.05, 0.1, and 0.2.

Higgins and Thompson [56] suggested a heterogeneity measure in meta-analysis, $I^2 = \tau^2/(\tau^2 + s^2)$, where s^2 is the 'typical' within-study variance,

$$s^2 = \frac{\sum_i w_i(k-1)}{\left(\sum_i w_i\right)^2 - \sum_i w_i^2},$$

and $w_i = 1/s_i^2$. In other words, I^2 describes the proportion of total variation across studies that is attributable to heterogeneity rather than within-study variability. The averages of s^2 over 1000 replicates are 0.09, 0.05, and 0.04 for $k = 5$, 10, and 30, respectively. Table II presents the average of I^2 over 1000 replicates for each combination of k and τ^2. As k increases, the average of the 'typical' within-study variance decreases, and I^2 increases with increasing k and τ^2.

The between-study variance τ^2 of the meta-analysis model is estimated using UMM, DL, ML, REML, and BM with a gamma$(2, 10^{-4})$ prior on τ as discussed. The overall effect μ is estimated by plugging in $\hat{\tau}^2$ from each estimation method for τ^2. The fixed-effects (FE) model is fitted by letting $\tau^2 = 0$. CIs for μ are estimated using Wald-type CIs on the basis of the expected information with FE, UMM, DL, ML and REML estimates, and Wald-type CIs based on the observed information with BM estimates. For the ML estimates, we also compare PL CIs and Wald-type CIs on the basis of the observed information.

For each simulated dataset, the parameters are also estimated by MCMC using WinBUGS 1.4.3 [57, 58] with a diffuse prior $N(0, 10^6)$ on μ and two different vague priors for the between-study variation: uniform(0,100) on τ and inverse-gamma$(10^{-3}, 10^{-3})$ on τ^2. After a 1000-iteration burn-in, μ and its 95% credible interval are estimated by the median and percentiles of 5000 posterior samples, respectively.

For the meta-regression model, we simulate data with different values of the between-study variance ($\tau^2 = 0$, 0.01, 0.05, and 0.1) for $k = 10$ and $m = 2$, and with the intercept and regression coefficients

Table II. Heterogeneity measure, I^2, for each simulation condition.

	k=5	k=10	k=30
τ^2=0	0	0	0
τ^2=0.01	0.16	0.19	0.22
τ^2=0.05	0.45	0.52	0.58
τ^2=0.1	0.60	0.68	0.73
τ^2=0.2	0.73	0.80	0.84

set to 1. Covariates are generated from a bivariate normal distribution with 0 means, unit variances, and a correlation of 0.3.

As in the meta-analysis model, we simulate 1000 replicates with heterogeneous within-study variances s_i^2. We use the same values of covariates over replications. The between-study variance τ^2 of the meta-regression model is estimated using ML, REML, and BM estimators.

The simulations are implemented using Stata, specifically, our own code is used for UMM, `metaan` is used for FE, DL, ML, REML, and PL, `gllamm` is used for BM, and `metareg` is used for ML and REML estimations of the model with covariates.

8.2. *Results*

8.2.1. Boundary estimates. First, we report the proportions of the estimates of τ that are on the boundary (less than 10^{-5}) when the true τ is not 0. BM yields no boundary estimates in any of the simulation conditions. As shown in Figure 5 (for τ up to $\sqrt{0.1}$), the proportions of boundary estimates for the other methods are large when k and τ are small (over 50% when $\tau = \sqrt{0.01}$ and $k = 5$) and decrease as k and τ increase. DL and REML perform similarly across all conditions and better than ML when $k = 5$ and $k = 10$. For $k = 30$, ML is close to DL and REML, and the proportion of boundary estimates is negligible when $\tau \geqslant \sqrt{0.05}$. For $k = 10$ and $k = 30$, UMM performs worst producing a substantial proportion of boundary estimates even for $k = 30$ when τ is large.

8.2.2. Log likelihood. To examine the decrease of the log likelihood from its maximum at $\hat{\tau}_{\mathrm{BM}}$, the likelihood-ratio test (LRT) statistics is calculated as

$$-2\left[l(\hat{\mu}_{\mathrm{BM}}, \hat{\tau}^2_{\mathrm{BM}}) - l(\hat{\mu}_{\mathrm{ML}}, \hat{\tau}^2_{\mathrm{ML}})\right]$$

for each replicate. For given values of $k = 5$, 10, and 30, the 0.99 quantile of the LRT statistic is largest when $\tau = 0$ with values 2.499, 1.766, and 1.159, respectively.

For testing $H_0 : \tau^2 = 0$, the asymptotic null distribution of the LRT statistic is $0.5\chi^2(0) + 0.5\chi^2(1)$ whose 0.99 quantile is 5.41. Crainiceanu and Ruppert [59] argue that this χ^2 mixture approximation can be very poor for finite samples and provide the exact LRT quantiles. Although our null value $\hat{\tau}_{\mathrm{BM}}$ is not 0 (but could be close to 0), we can compare our LRT statistics with Crainiceanu and Ruppert's [59] results for drawing a conservative conclusion, which shows that the 0.99 quantile of the exact distribution of the LRT statistics is 3.48 when $k = 5$ and study sizes are assumed to be infinite. This implies that BM estimates do not decrease the likelihood substantially from the maximum for most of the simulation conditions, and so the effect of the prior on the parameter estimates is sufficiently weak.

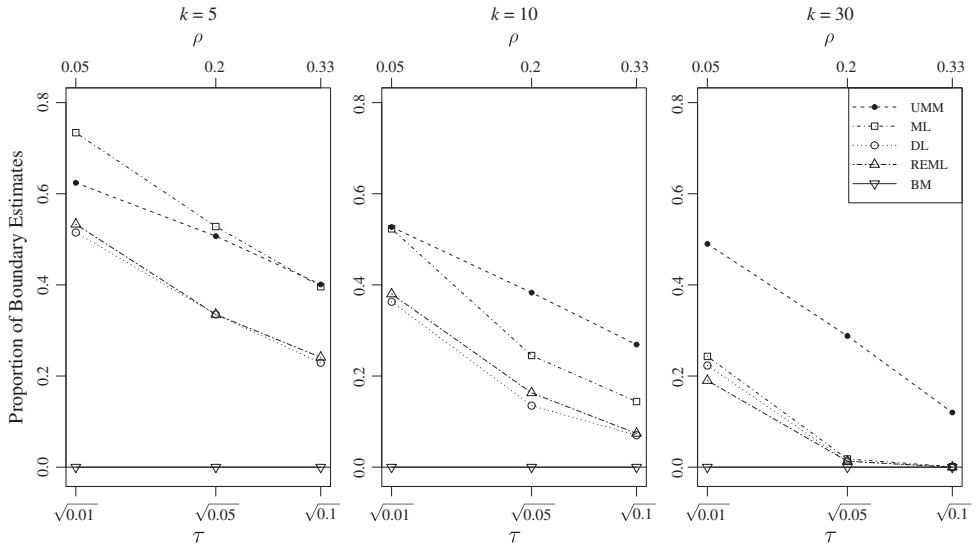

Figure 5. Proportions of the boundary estimates of τ for meta-analysis. DL, DerSimonian and Laird; UMM, unweighted method of moment; ML, maximum likelihood; REML, restricted maximum likelihood; BM, Bayes modal.

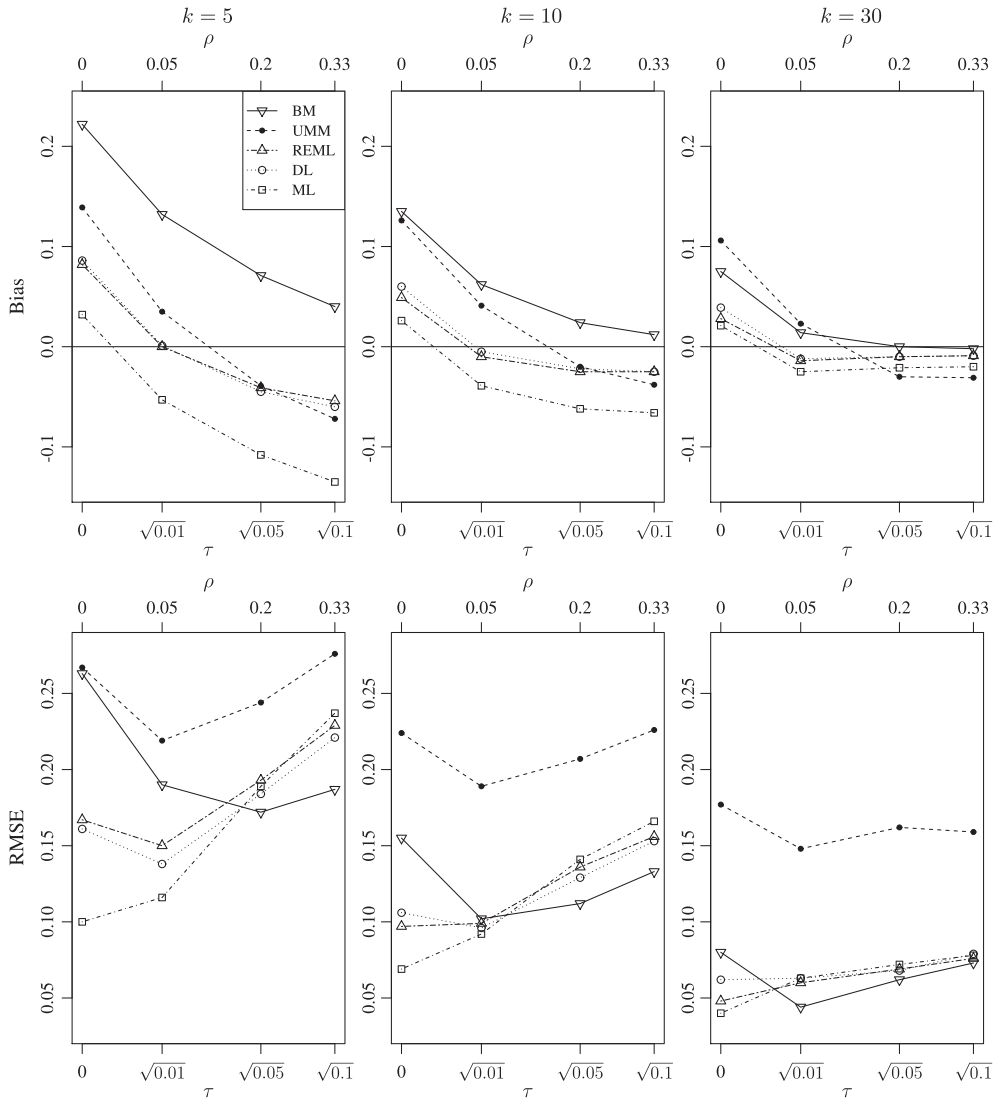

Figure 6. Estimated bias (top row) and root mean squared error (RMSE) (bottom row) of $\hat{\tau}$ for meta-analysis. DL, DerSimonian and Laird; UMM, unweighted method of moment; ML, maximum likelihood; REML, restricted maximum likelihood; BM, Bayes modal.

8.2.3. Estimation of τ. The top panels of Figure 6 summarize the estimated bias of the UMM, DL, ML, REML, and BM estimators for τ, with τ ranging only up to $\sqrt{0.1}$ for clarity. For all methods, the bias decreases as k increases for a given value of τ and is close to 0 for $k = 30$ when $\tau \geq \sqrt{0.01}$.

When $\tau = 0$, all estimation methods overestimate τ, but as τ increases, the bias decreases and rapidly becomes negative for DL, ML, and REML, with ML performing the worst. This problem of negative bias was also found by Sidik and Jonkman [60] and led them to conclude that the DL, ML, and REML estimators 'may only be recommended if the analyst is fairly certain that the heterogeneity in effect measures among studies is relatively small'. The BM estimator produces upward biased estimates across all conditions and performs better than ML and about as well as DL and REML when $\tau \geq \sqrt{0.05}$ (but with bias in the opposite direction).

The bottom panels of Figure 6 display the root mean squared error (RMSE) of $\hat{\tau}$. Consistent with the simulation results of Sidik and Jonkman [60], UMM yields larger RMSE than DL, ML, and REML, and we find that the RMSE for UMM is also larger than for BM even when $\tau = 0$. When $\tau = 0$, BM has larger RMSE than DL, ML, and REML but performs better than these estimators when τ is sufficiently large ($\tau \geq \sqrt{0.05}$ for $k = 5$ and $\tau \geq \sqrt{0.01}$ for $k = 10$). For $k = 30$, BM performs similarly to DL, ML, and REML.

8.2.4. Coverage probability of confidence intervals for μ. The estimated coverage probabilities of 95% CIs for μ estimated using UMM, DL, ML, REML, PL, and BM are presented in the top panels of Figure 7 (again, for τ up to $\sqrt{0.1}$). When $\tau = 0$, the coverage probabilities for all methods are above 0.95 because τ is overestimated by all methods in this case [11]. The coverage probabilities decrease with increasing τ for all k, particularly for small k. Consistent with other simulation studies [11, 61], DL, ML, and REML produce increasingly severe under-coverage as τ increases. As also found by Brockwell and Gordon [11], PL produces larger coverage and hence performs better than DL, ML, and REML unless τ is close to 0. The BM estimator produces even larger coverage and outperforms all methods when $\tau \geq \sqrt{0.05}$. For $\tau \geq \sqrt{0.01}$, the estimated coverage for BM is at least as close to 0.95 as for DL, ML, and REML and is substantially closer to 0.95 for larger values of τ.

In the bottom panels of Figure 7, we also compare the estimated coverage probabilities of two Wald-type CIs for ML: one based on the expected information (ML: exp), which is also shown in the top panels, and the other based on the observed information (ML: obs) with the profile likelihood CI. As expected, the estimated coverage probabilities for ML based on the observed information are slightly higher than the ones based on the expected information. However, both Wald-type CIs perform worse than PL.

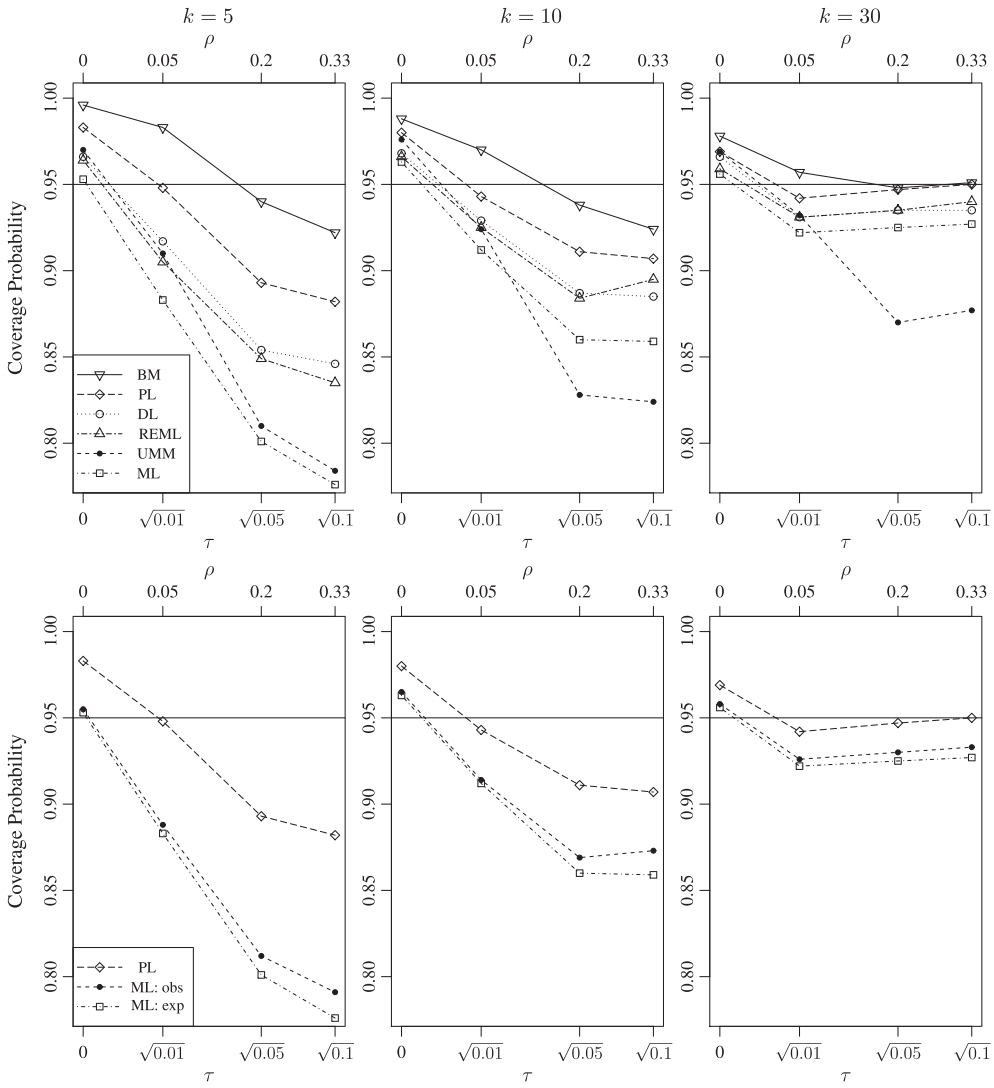

Figure 7. Estimated coverage probabilities for the UMM, DL, ML, REML, PL, and BM estimators (top row), and ML and PL estimators (bottom row) of 95% confidence intervals for μ for meta-analysis. The nominal level, 0.95, is shown as a horizontal solid line. DL, DerSimonian and Laird; UMM, unweighted method of moment; ML, maximum likelihood; REML, restricted maximum likelihood; BM, Bayes modal; PL, profile likelihood.

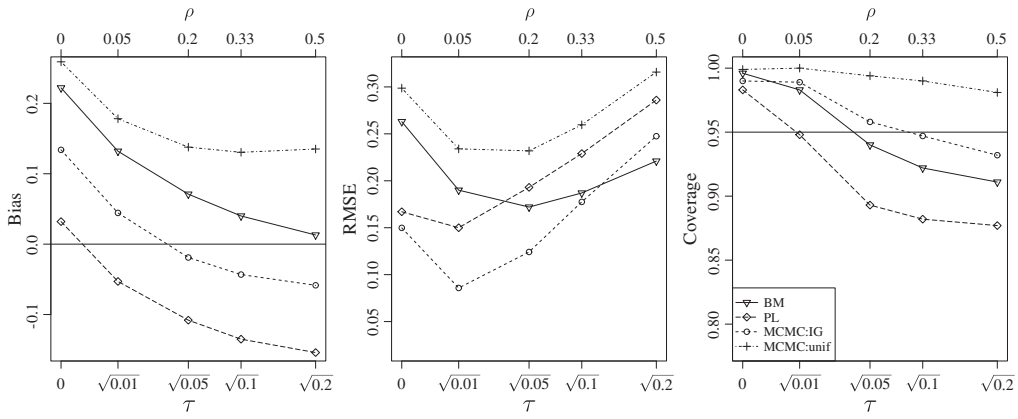

Figure 8. Estimated bias and root mean squared error (RMSE) of τ and coverage of 95% credible intervals for μ by posterior median estimation using MCMC. ($k = 5$). BM, Bayes modal; PL, profile likelihood.

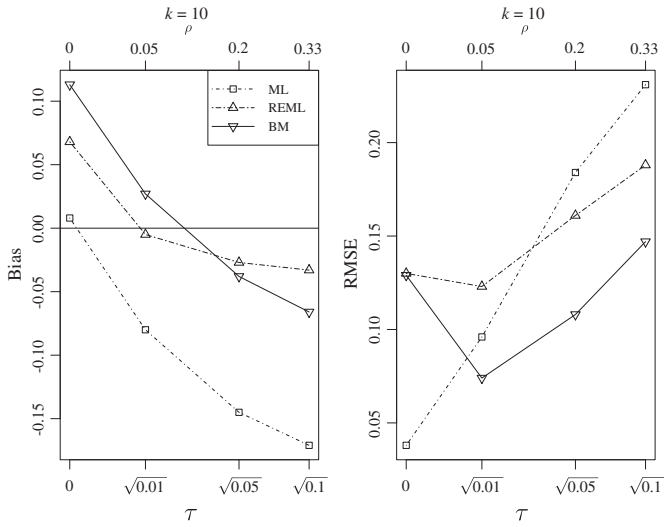

Figure 9. Estimated bias and root mean squared error (RMSE) of τ in a meta-regression model with two covariates. ML, maximum likelihood; REML, restricted maximum likelihood; BM, Bayes modal.

For constructing Wald CIs for μ, Berkey *et al.* [10] suggest using $t_{1-\alpha/2}(k-4)$ for the multiplier of the standard error instead of the normal quantile. In the context of testing, Hartung and Knapp [62] use the t distribution with $k-1$ degrees of freedom. We also provide coverage probabilities with t quantiles with $k-1$ and $k-4$ degrees of freedom in the supplementary material.[‡] When $k = 5$ and 10, the Wald CI with traditional estimators for τ^2 tends to have better coverage than those shown in Figure 7 when $t_{1-\alpha/2}(k-1)$ is used for the multiplier. The CI with $t_{1-\alpha/2}(k-4)$ tends to produce over-coverage, especially for $k = 5$, where coverage is 100% for all values of μ and τ considered in the simulation.

8.2.5. Results for MCMC. The estimated bias and RMSE of τ for $k = 5$ are displayed in the left and middle panels of Figure 8, respectively (for τ up to $\sqrt{0.2}$). The uniform prior produces the largest bias and RMSE regardless of the value of τ. The estimates of τ using the inverse-gamma prior tend to be smaller than the BM estimates, with smaller bias for small τ but larger (downward bias) for larger τ. The RMSE of the inverse gamma is the smallest for $\tau < \sqrt{0.05}$ but larger than or similar to the RMSE of BM for $\tau \geq \sqrt{0.1}$. The right panel of Figure 8 compares the coverage of 95% credible intervals for μ estimated by MCMC with the coverage of 95% CIs estimated by BM and PL for $k = 5$. Both the uniform and the inverse-gamma priors produce higher coverage than BM and PL. For the uniform prior,

[‡]*Supporting information may be found in the online version of this article.*

coverage is substantially larger than 0.95 even for $\tau = \sqrt{0.2}$ where all the other methods have lower than nominal coverages. For $k = 10$ and 30 (supplementary materials), the coverage, bias, and RMSE follow a similar pattern, but the difference between estimators fades as k increases.

8.2.6. Meta-regression. Figure 9 shows the estimated bias and RMSE of τ for the meta-regression model with two covariates for $k = 10$, estimated using ML, REML, and BM. The results are similar as for the meta-analysis model without covariates (Figure 6, middle panels), except that BM has negative bias for large τ. As before, BM performs best in terms of RMSE when $\tau \geq \sqrt{0.01}$.

9. Discussion

When unexplained between-study heterogeneity is believed to exist, as is the case in most meta-analyses, random-effects meta-analysis or meta-regression is more appropriate than the fixed-effects counterparts. However, estimates of the between-study (residual) variance are imprecise when the number of studies is small and are often on the boundary of 0, yielding the unrealistic fixed-effects results. As seen with the antiplatelet therapy data example, the between-study variance estimate can have a greater impact on inferences when the studies vary more in their standard errors.

We proposed using a BM estimator with a gamma prior for the between-study standard deviation. Because this estimator never produces boundary estimates, the meta-analysis model can include unexplained heterogeneity even when the data do not provide enough information for the between-study variation. At the same time, our method still respects the data in the sense that it leads to small decreases in the log likelihood. Therefore, our approach is preferable to traditional estimators, such as ML or restricted ML estimators, when boundary estimates occur.

Compared with full Bayes estimation, our approach does not require simulation with convergence checking and is simple to use with the `gllamm` command in Stata. In addition, it is computationally as efficient as ML estimation and, in fact, potentially more efficient as it avoids the slow convergence that can occur if the ML estimate is on the boundary. The posterior median estimator with the inverse-gamma$(10^{-3}, 10^{-3})$ prior for the between-study variance performs quite well, but not as well as the BM estimator when moderate between-study variation exists.

When the traditional estimators do not produce boundary estimates, our approach tends to be a little more conservative by yielding larger variance estimates and wider estimated CIs for the mean effect size. As the number of studies increases, our estimator approaches the traditional estimators.

Acknowledgements

Contract/grant sponsor: The research reported here was supported by the Institute of Education Sciences; contract/grant number: (R305D100017)

References

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(3):177–188.
2. National Research Council. *Combining Information: Statistical Issues and Opportunities for Research*. National Academic Press: Washington, DC, 1992.
3. Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). *Annual Review of Public Health* 1996; **17**(1):1–23.
4. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A* 2009; **172**(1):137–159.
5. Borenstein M, Hedges L, Higgins J, Rothstein H. *Introduction to Meta-analysis*. Wiley: Chichester, 2009.
6. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 1994; **309**(6965):1351–1355.
7. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Statistics in Medicine* 2002; **21**(11):1503–1511.
8. Raudenbush SW, Bryk AS. Empirical Bayes meta-analysis. *Journal of Educational Statistics* 1985; **10**(2):75–98.
9. Hedges LV. Commentary. *Statistics in Medicine* 1987; **6**(3):381–385.
10. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**(4):395–411.
11. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001; **20**(6):825–840.

12. Sánchez-Meca J, Marín-Martínez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* 2008; **13**(1):31–48.
13. Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B* 1995; **57**(1):45–97.
14. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics in Medicine* 2000; **19**(13):1707–1728.
15. Curcio D, Verde P. Comment on: Efficacy and safety of tigecycline: A systematic review and meta-analysis. *Journal of Antimicrobial Chemotherapy* 2011; **66**(12):2893–2895.
16. Hardy R, Thompson S. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 1998; **17**(8):841–856.
17. Overton R. A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods* 1998; **3**(3):354.
18. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* 2005; **30**(3):261–293.
19. Higgins JPT, Whitehead A. Borrowing strength from external trials in meta-analysis. *Statistics in Medicine* 1996; **15**(24):2733–2749.
20. Longford NT. On estimating standard errors in multilevel analysis. *Journal of the Royal Statistical Society, Series D* 2000; **49**(3):389–398.
21. Chung Y, Rabe-Hesketh S, Dorie V, Gelman A, Liu J. A non-degenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika* 2013:In press. DOI: 10.1007/s11336-013-9328-2.
22. Lawlor D, Hopker S. The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *British Medical Journal* 2001; **322**:1–8.
23. Higgins J, Thompson S. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* 2004; **23**(11):1663–1682.
24. Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika* 1971; **58**(3):545–554.
25. Kontopantelis E, Reeves D. metaan: Random-effects meta-analysis. *The Stata Journal* 2010; **10**(3):395–407.
26. Harbord R, Higgins JPT. Meta-regression in Stata. *The Stata Journal* 2008; **8**(4):493–519.
27. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002; **324**(7329):71–86.
28. Hedges LV. A random effects model for effect sizes. *Psychological Bulletin* 1983; **93**(2):388–395.
29. Rao C. *Linear Statistical Inference and Its Applications*, (Second Edition), Wiley: New York, 1973.
30. Demidenko E. *Mixed Models: Theory and Applications*. Wiley: Hoboken, NJ, 2004.
31. Morris C. Parametric empirical Bayes inference—theory and applications. *Journal of the American Statistical Association* 1983; **78**(381):47–55.
32. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 2010; **36**(10):1–48.
33. Hardy R, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**(6):619–629.
34. Efron B, Hinkley D. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* 1978; **65**(3):457–483.
35. Fisher R. Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* 1925; **22**(5):700–725.
36. Pawitan Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press: Oxford, 2001.
37. Lambert P, Sutton A, Burton P, Abrams K, Jones D. How vague is vague? a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in Medicine* 2005; **24**(15):2401–2428.
38. Pauler DK, Wakefield JC. Modeling and implementation issues in Bayesian meta-analysis. In *Meta-Analysis in Medicine and Health Policy*, Stangl DK, Berry DA (eds). Marcel Dekker, Inc: New York, 2000.
39. Abrams K, Sanso B. Approximate Bayesian inference for random effects meta-analysis. *Statistics in Medicine* 1998; **17**(2):201–218.
40. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**(24):2733–2749.
41. Sutton A, Abrams K. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**(4):277–303.
42. Spiegelhalter D, Abrams K, Myles J. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley: New York, 2004.
43. Galindo-Garre F, Vermunt J, Bergsma W. Bayesian posterior mode estimation of logit parameters with small samples. *Sociological Methods & Research* 2004; **33**(1):88–117.
44. Maris E. Estimating multiple classification latent class models. *Psychometrika* 1999; **64**(2):187–212.
45. Galindo-Garre F, Vermunt J. Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika* 2006; **33**(1):43–59.
46. Swaminathan H, Gifford JA. Bayesian estimation in the two-parameter logistic model. *Psychometrika* 1985; **50**(3):349–364.
47. Mislevy RJ. Bayes modal estimation in item response models. *Psychometrika* 1986; **51**(2):177–195.
48. Tsutakawa RK, Lin HY. Bayesian estimation of item response curves. *Psychometrika* 1986; **51**(2):251–267.
49. Berger J. The case for objective Bayesian analysis. *Bayesian Analysis* 2006; **1**(3):385–402.

50. Fay RE, Herriot RA. Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 1979; **74**(366):269–277.
51. Morris C. Mixed model prediction and small area estimation (with discussions). *Test* 2006; **15**(1):72–76.
52. Li H, Lahiri P. An adjusted maximum likelihood method for solving small area estimation problems. *Journal of Multivariate Analysis* 2010; **101**(4):882–892.
53. Morris C, Tang R. Estimating random effects via adjustment for density maximization. *Statistical Science* 2011; **26**(2):271–287.
54. Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 2005; **128**(2):301–323.
55. Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modeling Using Stata*, (Third Edition), Stata Press: College Station, 2012.
56. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**(11):1539–1558.
57. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000; **10**:325–337.
58. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The Bugs Book*. Chapman & Hall/CRC Press: Boca Raton, FL, 2012.
59. Crainiceanu C, Ruppert D. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B* 2004; **66**(1):165–185.
60. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 2007; **26**(9):1964–1981.
61. Sánchez-Meca J, Marín-Martínez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* 2008; **13**(1):31–48.
62. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine* 2001; **20**(12):1771–1782.